

Prädiktive Analytik aus der Perspektive der Klinischen Psychologie und Psychotherapie

Kevin Hilbert · Ulrike Lueken

Institut für Psychologie, Humboldt-Universität zu Berlin, Berlin, Germany

Schlüsselwörter

Einzelfallprädiktion · Maschinelles Lernen · Kognitive Verhaltenstherapie · Biomarker · Prognose

Zusammenfassung

Während in der somatischen Medizin mittlerweile eine Vielzahl von biologischen Markern für die Diagnostik und Therapieplanung vorliegen, gibt es keine vergleichbaren biologischen oder psychologischen Marker für psychische Störungen. Hier sind die Pathogenese und Wirkung psychotherapeutischer Interventionen durch eine Vielzahl miteinander interagierender Faktoren determiniert. Die prädiktive Analytik verfügt mit dem maschinellen Lernen über eine aussichtsreiche Methode, komplexe Muster und Interaktionen zwischen verschiedenen Variablen in Aussagen für den individuellen Patienten zu übersetzen. Diese Methoden bestimmen ("lernen") aus bereits vorhandenen Daten die Beziehung zwischen Prädiktoren und Ergebnissen und können anschließend das entwickelte Modell auf neue Daten, bei denen das Ergebnis noch offen ist, anwenden. Zuvor muss aber zwingend geprüft werden, ob das Gelernte tatsächlich bedeutungsvoll ist. Zur Illustration des Ansatzes stellen wir eine Reihe von Studien vor, die das Paradigma der prädiktiven Analytik für diagnostische Fragestellungen, Vorhersage von Risikoverläufen sowie zur Prognose von Psychotherapieergebnissen genutzt haben. Die Ergebnisse sind vielversprechend; vor einem Einsatz in der klinischen Praxis muss die Vorhersagegenauigkeit jedoch weiter gesteigert und in verschiedenen Settings und Populationen überprüft werden. Zur Verbesserung der Vorhersagegüte scheinen insbesondere die Berücksichtigung unterschiedlicher

Datenmodalitäten wie klinische Maße, (f)MRT-Daten und genetische Daten sowie der Fokus auf Variablen, die Mechanismen von Psychopathologie und Veränderungsmechanismen gut abbilden, sinnvoll. Darüber hinaus sollte eine enge Zusammenarbeit mit Vertretern von Praktikern und Betroffenen stattfinden, um die Akzeptanz solcher Marker zu gewährleisten. Wenn dies gelingt, bieten derartige Marker das Potenzial, die Diagnosesicherheit insbesondere in schwierigen Fällen deutlich zu erhöhen, mögliche Risikoverläufe früh zu identifizieren und die Zuweisung von Patienten zu den für sie bestmöglichen Behandlungen zu unterstützen.

© 2020 S. Karger AG, Basel

Predictive Analytics from a Mental Health Perspective

Key Words

Single case prediction · Machine learning · Cognitive behavioral therapy · Biomarker · Prognosis

Abstract

While there is a plethora of biomarkers for diagnosis and treatment selection in somatic medicine, no comparable biological or psychological markers are available in mental health. Following the bio-psycho-social model, both the pathogenesis and treatment effects are determined by many concurring factors. With machine learning, predictive analytics offer a promising set of tools for translating patterns and interactions in and between a variety of

variables into a conclusion for the individual patient. These methods “learn” the association between predictors and outcomes from already available data and can then apply the resulting model on new data, for which the outcome is still open. However, it is crucial to evaluate beforehand whether the learned model is meaningful. To illustrate this approach, we present a number of studies that used predictive analytics for diagnostics, for predicting risk trajectories and for predicting psychotherapy treatment outcomes (“theranostics”). Their results are promising, but prior to clinical practice the prediction accuracy has to be increased and tested in different settings and populations. For increasing prediction performance, combining several data modalities, such as clinical, neurostructural, -functional and genetic data, and focusing on variables that map mechanisms of psychopathology and change, are reasonable. Moreover, teaming up with clinician and patient representatives is recommended for increasing the acceptance of such markers and discussing the ethical and societal implications of predictive analytics in mental health. If successful, predictive analytics bear the potential to increase diagnostic reliability particularly in challenging cases, to identify potentially negative trajectories early on and to support allocating patients to their individually optimal treatment.

© 2020 S. Karger AG, Basel

Warum prädiktive Analytik?

Klinische Entscheidungen in der somatischen Medizin werden mittlerweile von einer Vielzahl von Indikatoren – biologischen Markern (siehe Kasten 1) – unterstützt [Strimbu und Tavel, 2010]. Über den Symptombereich des Patienten hinaus liefern diese Biomarker wesentliche Informationen zum Risiko zukünftiger Krankheitsgenese, zur (Differential-)Diagnostik und zur Indikation verschiedener Behandlungsoptionen. Vergleichbare Marker für unser Feld der psychischen Störungen liegen bisher nicht vor [Venkatasubramanian und Keshavan, 2016]. Dies liegt maßgeblich daran, dass sowohl die Entstehung [Zuckerman, 1999] als auch der Verlauf [Green et al., 2013] psychischer Störungen sowie die Wirkmechanismen psychotherapeutischer Interventionen [Wampold und Imel, 2015] durch eine Vielzahl sich wechselseitig beeinflussender Faktoren determiniert sind, wie es im bio-psycho-sozialen Modell deutlich wird [Engel, 1977]. Gelänge es trotz dieser multifaktoriellen Determiniertheit, vergleichbare Marker für psychische Störungen zu entwickeln, erlaubten diese die Abschätzung individueller Verläufe sowie die Adaption etablierter Interventionen an den einzelnen Patienten. Dies ist der Ansatzpunkt der personalisierten Medizin [Ozomaro et al., 2013]. Von ihr erhofft man sich durch die bessere

Kasten 1. Was ist ein Biomarker?

Biologische Marker (kurz: Biomarker) sind objektiv messbare biologische Indikatoren für andere, entweder zurzeit noch nicht messbare oder erst in der Zukunft auftretende Prozesse oder Ereignisse. Die beiden zentralen Definitionen liegen leider nur auf Englisch vor: “A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention” [Biomarkers Definitions Working Group, 2001, p. 91] sowie von der Weltgesundheitsorganisation mit “Any substance, structure, or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease” [World Health Organization and International Programme on Chemical Safety, 2001]. Auch für die psychische Gesundheit werden Kandidaten für biologische Marker wie beispielsweise genetische Polymorphismen oder neuronale Aktivierungsmuster diskutiert. Grundsätzlich wären für die prädiktive Analytik jedoch auch intrapsychische oder soziale Marker denkbar, solange diese reliabel zu erfassen und kriteriumsvalid hinsichtlich der interessierenden abhängigen Variable sind. Im vorliegenden Manuskript sprechen wir von einem Marker, wenn eine einzelne Klassifikation oder Prädiktion erfolgt, auch wenn dafür auf mehrere Prädiktorvariablen zurückgegriffen wird (zusammengesetzter Marker).

Passung von Patient und Intervention insgesamt effizientere und effektivere Behandlungen, also sowohl individuelles Leiden und Beeinträchtigung zu reduzieren, als auch auf gesellschaftlicher Ebene Kosten zu sparen bzw. mit den verfügbaren Mitteln bessere Ergebnisse zu erzielen. Mithilfe der prädiktiven Analytik sollen entsprechend leistungsfähige Modelle entwickelt werden, die multifaktorielle Informationen integrieren und so Marker für die genannten Ziele bereitstellen. Ein Kernstück der prädiktiven Analytik ist das maschinelle Lernen. Maschinelle Lernalgorithmen im Bereich der somatischen Medizin waren beispielsweise klinischen Experten hinsichtlich des Erkennens von erblich bedingten Syndromen [Gurovich et al., 2019] oder der Wahl der optimalen Behandlung bei Sepsis überlegen [Komorowski et al., 2018]. Ein komplexes Deep-Learning Modell absolvierte 2017 den Mediziner-Test in China erfolgreich und erzielte eine höhere Punktzahl als 96,3% der menschlichen Teilnehmer [Wu et al., 2018].

Maschinelles Lernen als aussichtsreiche Methode zur Einzelfallklassifikation und -prädiktion

Maschinelles Lernen ist ein Oberbegriff für eine Vielzahl von Verfahren, die insbesondere eine gemeinsame Herangehensweise an ein Problem eint. Letztlich geht es darum, aus vorliegenden Daten bestimmte Merkmalskombinationen (Muster) herauszufiltern [Bishop, 2006]. Die Daten sind meist hochdimensional, und Analysen fin-

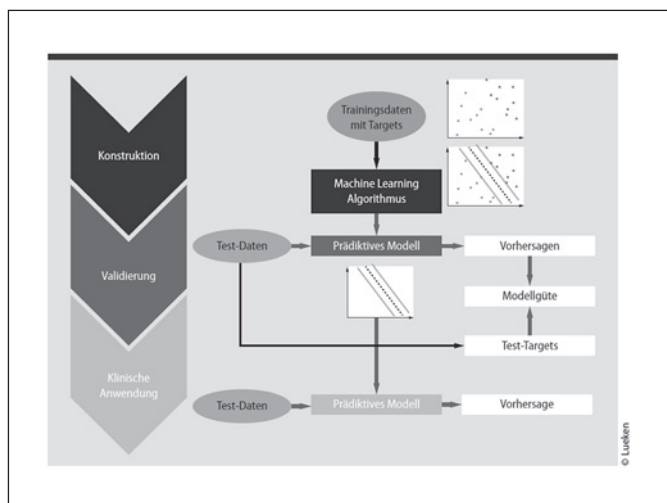


Abb. 1. (Modifiziert nach Lueken und Hahn [2017]). Schematische Abbildung der verschiedenen Schritte beim maschinellen Lernen. Mit bekannten Daten wird während der Trainingsphase das Modell konstruiert und anschließend in der Testphase evaluiert. Dazu werden die Voraussagen des Modells mit den tatsächlichen Daten verglichen und die Güte der Übereinstimmung über verschiedene Indizes, wie beispielsweise die Genauigkeit (zwischen 0 und 100% Übereinstimmung), Sensitivität, Spezifität oder ROC-Kurven, quantifiziert. Ist das Modell gut genug für den Einsatz in der Praxis, kann es nun für Daten eingesetzt werden, bei denen das interessierende Ergebnis tatsächlich noch nicht bekannt ist.

den bisweilen in sehr großen, mitunter unüberschaubaren Datensätzen statt. Daher werden diese Ansätze auch häufig mit dem Schlagwort “big data” in Zusammenhang gebracht. Beim *supervidierten* maschinellen Lernen, um das es im Weiteren ausschließlich gehen wird, versucht ein statistischer Algorithmus in vorliegenden Daten ein Muster zu finden, dass bestmöglich zwischen verschiedenen Gruppen oder Klassen unterscheidet [Bishop, 2006; Kotsiantis et al., 2006; Kotsiantis, 2007]. Welche Gruppen es dabei gibt, wird vom Entwickler vorgegeben. Ein Beispiel wäre hier, anhand der Antworten in einigen Symptomfragebögen nach einem Muster zu suchen, das Auskunft darüber gibt, ob ein Patient zur Klasse “wird nach Therapieende remittieren” oder zur Klasse “wird nach Therapieende nicht remittieren” gehört. Der Algorithmus “lernt” beim maschinellen Lernen insofern, als er mit zunehmender Anzahl an gesehenen Patienten bestimmte Merkmale, die sich gut zur Unterscheidung beider Klassen eignen, höher gewichtet und kombiniert, während schlecht geeignete Merkmale nach und nach immer weniger beachtet werden. Insofern “lernt” der Algorithmus sukzessive zwischen den genannten Klassen zu unterscheiden. Die Klassen müssen dabei nicht kategorial sein, auch dimensionale Daten können vorhergesagt werden [Mohri et al., 2018]. Im Gegensatz dazu sind beim *unsupervidierten* maschinellen Lernen keine Klassen vorgegeben, der Algorithmus versucht also selbst, in den Daten eine Struktur

zu finden [Bishop, 2006]. Ein Beispiel wäre hier, nach Gruppen von Patienten zu suchen, die sich in ihren Antworten in einigen Symptomfragebögen sehr ähnlich sind. Clusteranalysen sind typische Vertreter dieser Gruppe von Verfahren. Unsupervidiertes maschinelles Lernen bietet großes Potenzial für die Identifikation neuer, klinisch relevanter Endophenotypen. Im Folgenden beschränken wir uns jedoch auf die Unterscheidung vorgegebener Gruppen mittels supervidiertem Lernen.

Zentral beim maschinellen Lernen ist, dass es sich um ein schrittweises Verfahren handelt (Abb. 1). In einem ersten Schritt lernt der Algorithmus anhand vorgegebener Daten, daher wird dieser Schritt häufig als Training bezeichnet [Bishop, 2006; Mohri et al., 2018]. In einem zweiten Schritt muss nun überprüft werden, ob das gelernte Muster tatsächlich bedeutungsvoll ist. Dieser Schritt wird entsprechend auch als Test bezeichnet. Hier wird der Algorithmus genutzt, um anhand der gelernten Muster Voraussagen für bisher nicht genutzte Daten zu machen [Bishop, 2006; Mohri et al., 2018]. Indem die Voraussagen mit den tatsächlichen Daten verglichen werden, kann die Güte des gelernten Modells eingeschätzt werden [vgl. Sokolova und Lapalme, 2009]. Ein gutes Modell wird auch auf bisher nicht genutzten Daten Voraussagen produzieren, die häufig mit der Wirklichkeit übereinstimmen, während ein schlechtes Modell dies eher nicht tun wird. Es ist zwingend notwendig, die Testphase auch tatsächlich mit neuen Daten durchzuführen, da andernfalls die Modellgüte deutlich überschätzt wird und keine unabhängige Vorhersage erzeugt wird. In der dritten Phase, der Anwendungsphase, kann der Algorithmus nun genutzt werden, um Voraussagen an Daten mit tatsächlich unbekannten abhängigen Variablen in der Praxis zu generieren. Die Anwendung auf den individuellen Einzelfall, also idealerweise auf den nächsten Patienten in der klinischen Praxis, ist hier also im Gegensatz zu konventionellen Analysen von vornherein als Ziel bedacht [siehe Kasten 2 und vgl. Yarkoni und Westfall, 2017].

Ansätze des maschinellen Lernens bieten für die prädiktive Analytik eine Reihe von Vorteilen: (1) Sie sind besonders geeignet, eine Vielzahl von möglichen Prädiktorvariablen zusammen zu betrachten [Bzdok et al., 2018]. Während einfache Algorithmen mitunter nur rein additive oder lineare Zusammenhänge modellieren, können andere Verfahren auch komplexere Interaktionen abbilden [vgl. Wu et al., 2008]. Der Vorteil des maschinellen Lernens wird dabei meist umso größer, je mehr Daten zur Verfügung stehen (vgl. “big data”). Prinzipiell lässt sich die Rolle einzelner Prädiktorvariablen im Modell vergleichen, allerdings nimmt die Interpretierbarkeit gerade in komplexeren Modellen meist schnell ab. (2) Zudem muss kein explizites Modell vorgegeben werden, nach dem der Prädiktionsalgorithmus lernt [Bzdok et al., 2017]. Zentral für das Paradigma der prädiktiven Analy-

Kasten 2. Von der Regression zum maschinellen Lernen

Bei den Modellen des maschinellen Lernens handelt es sich nicht notwendigerweise nur um neue Verfahren, die sich von in der Psychologie bereits bekannten statistischen Modellen prinzipiell unterscheiden. Auch Standardverfahren wie lineare oder logistische Regressionen können im Sinne des maschinellen Lernens eingesetzt werden; tatsächlich sind Regressionsverfahren in vielen Softwarepaketen zum maschinellen Lernen enthalten, wie beispielsweise in scikit-learn [Pedregosa et al., 2011]. Der Unterschied liegt eher in der Art der Anwendung begründet, wie der Artikel von Yarkoni und Westfall [2017] im Detail darlegt. Zentral ist, dass Regressionsmodelle in der Psychologie traditionell nicht zur Vorhersage in einem unabhängigen Datensatz genutzt werden. In der Sprache des maschinellen Lernens handelt es sich also um das Trainieren eines Modells ohne eine entsprechende unabhängige Testung. Dies birgt jedoch die Gefahr der Überschätzung der Generalisierbarkeit und Genauigkeit des Modells. Prinzipiell könnten klassische Regressionsmodelle aber zur Testung in einem unabhängigen Datensatz genutzt werden.

Zudem bleiben lineare und logistische Regressionen vergleichsweise simple Modelle. Komplexere Algorithmen zusammen mit der in den letzten Dekaden exponentiell gewachsenen Rechenleistung erlauben nun aber auch die Modellierung nicht-linearer Zusammenhänge. Obwohl die Verwendung einfacher Modelle aus Praktikabilitätsicht attraktiv erscheint, werden aufgrund der multifaktoriellen Determiniertheit des bio-psycho-sozialen Modells möglicherweise nur komplexere Modelle die notwendige Genauigkeit für einen praktischen Einsatz erreichen.

tik ist aber (3) die Fähigkeit, die verschiedenen Prädiktorvariablen derart zu einer Entscheidungsfunktion zu verknüpfen, dass für jede Variablenkombination eine eindeutige Gruppenzuordnung erzeugt werden kann. Sie ermöglichen also eine Aussage über einen einzelnen Patienten mit einer individuellen Merkmalsausprägung [Orrù et al., 2012].

Viele Studien zur Klassifikation und Prädiktion mittels maschinellem Lernen nutzen bildgebende Daten, und dort insbesondere Magnetresonanztomographie (MRT)-Aufnahmen [z.B. Gong et al., 2014b; Lueken et al., 2014; Hahn et al., 2015; Whitfield-Gabrieli et al., 2016; Hilbert et al., 2017; Reggente et al., 2018]. Gerade für den Bereich der Prädiktion von Therapieoutcomes gibt es zudem eine beträchtliche Anzahl an Arbeiten, die soziodemographische oder klinische Routinedaten nutzten [z.B. Askland et al., 2015; Iniesta et al., 2016; Deisenhofer et al., 2018; Hilbert et al., 2020; Cohen et al., 2020]. Aber auch für andere Datenmodalitäten, wie beispielsweise Ecological Momentary Assessment (EMA), liegen bereits Untersuchungen zur Einzelfallprädiktion vor [van Breda et al., 2018]. Die multimodale Kombination verschiedener Datenebenen findet sich bislang aber noch selten [z.B. Hilbert et al., 2017], was an der begrenzten Verfügbarkeit entsprechend aufwendiger, insbesondere longitudinaler

Studien liegen mag. Daher ist noch unklar, welche Marker oder Marker-Kombinationen eine besonders genaue Prädiktion erlauben würden, oder, aus einer gesundheitsökonomischen Perspektive, welche Marker durch näherungsweise gleiche, kosteneffektivere Merkmale substituiert werden könnten. In anderen Worten: für welche diagnostische oder prognostische Fragestellung würde sich ein MRT-Scan rechnen bzw. wann ließen sich vergleichbare Prädiktionsgenauigkeiten auch mit kostengünstigeren (z.B. Fragebogen-)Daten erzielen?

Ausgewählte Studien zur prädiktiven Analytik im Bereich der psychischen Gesundheit

Bisherige Studien zur prädiktiven Analytik im Bereich der psychischen Gesundheit lassen sich besonders drei Bereichen zuordnen: (1) klassifikatorische Studien, in denen Marker gesucht werden, die bestimmte Störungen oder klinische Zustände voneinander abgrenzen. Beispiele für entsprechende Fragestellungen wären hier: Liegt bei einem älteren Patienten mit kognitiven Defiziten eine leichte kognitive Beeinträchtigung oder eine Alzheimererkrankung vor? Liegt bei einem Patienten mit einer depressiven Episode eine unipolare oder eine bipolare Störung vor? (2) Prädiktionsstudien, bei denen Risikoverläufe vorhergesagt werden sollen. Als Beispiel wäre hier die Frage zu nennen, welche Personen mit einem Trauma nach einer Naturkatastrophe eine vollausgeprägte Posttraumatische Belastungsstörung (PTBS) entwickeln. (3) Prädiktionsstudien, bei denen Therapieoutcomes vorhergesagt werden sollen. Beispiele wären hier: Wird ein Patient nach einer kognitiven Verhaltenstherapie (KVT) remittieren? Wird eine Patientin von einer KVT oder einer psychodynamischen Behandlung mehr profitieren? Im Folgenden sollen einige beispielhafte Studien genauer vorgestellt werden.

Differentialdiagnostische Marker wurden für eine Reihe von Störungen untersucht, beispielsweise zur Abgrenzung von verschiedenen Subtypen von Phobien [Lueken et al., 2014], zur Unterscheidung depressiver Probanden von gesunden Kontrollprobanden [Hahn et al., 2011] oder zur Klassifikation schizophrener Patienten von Gesunden [Shim et al., 2016]. Einen Mehrwert könnten solche Marker jedoch besonders für Störungen bieten, die mit aktuellen Methoden wie klinischen Interviews oder Symptomfragebögen mitunter schwer von anderen Störungen zu unterscheiden sind. So ist beispielsweise die Generalisierte Angststörung (GAS) von einer unipolaren depressiven Störung in der Praxis mitunter schwer zu differenzieren. Bei beiden können repetitive, negative Denkmuster im Vordergrund stehen und mit Begleitscheinungen wie leichter Ermüdung, Schlafstörungen, Konzentrationsproblemen, Reizbarkeit oder Muskelschmerzen einhergehen

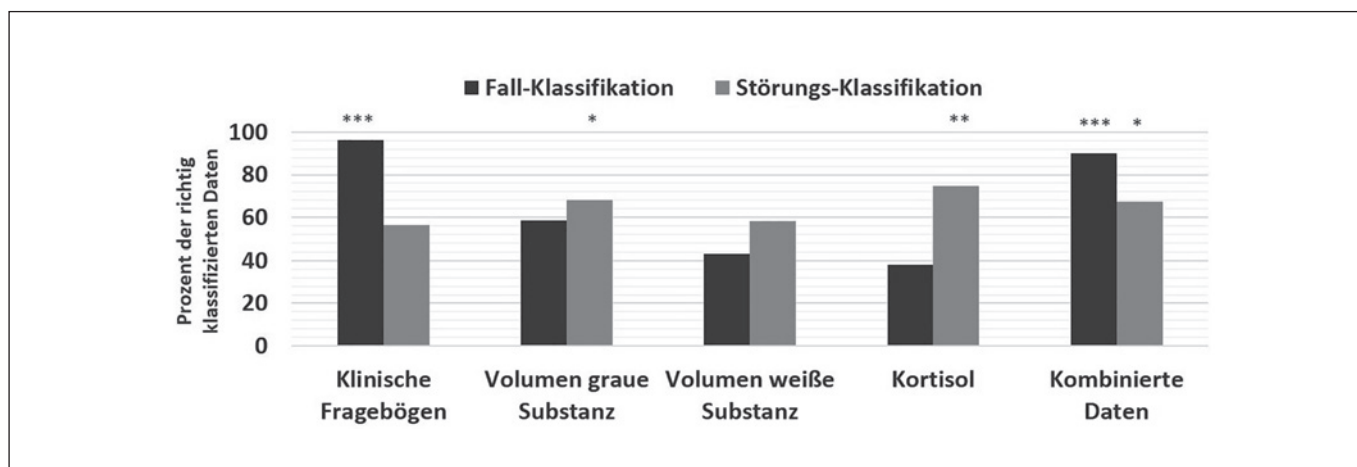


Abb. 2. (Nach Hilbert et al. [2017] © 2016 The Authors. *Brain and Behavior* published by Wiley Periodicals, Inc., made available under the terms of the Creative Commons Attribution License [CC BY; <https://creativecommons.org/licenses/by/3.0/>]). Ergebnisse der Klassifikationsstudie von Hilbert et al. [2017]. Zur Klassifikation wurden die einzelnen Datenmodalitäten separat und zuletzt

in Kombination genutzt. Fall-Klassifikation: Unterscheidung der Patienten mit einer Störung von den gesunden Kontrollprobanden. Störungs-Klassifikation: Unterscheidung der Patienten mit einer GAS (mit und ohne komorbide unipolare depressive Störung) von denjenigen ohne GAS. *** $p < 0,001$, ** $p < 0,01$, * $p < 0,05$.

[APA, 2000, 2013]. Die diagnostischen Kriterien der Erkrankungen überlappen zudem deutlich [Zbozinek et al., 2012] und beide Störungen sind hochgradig komorbid [Carter et al., 2001; Kessler et al., 2005]. Für die primärärztliche Versorgung liegen zudem empirische Daten vor, die zeigen, dass die GAS häufig nicht erkannt oder fehldiagnostiziert wird [Wittchen et al., 2002; Munk-Jørgensen et al., 2006; Calleo et al., 2009; Vermani et al., 2011]. In einer Untersuchung an gesunden Kontrollprobanden, GAS Probanden mit und ohne komorbide depressive Störung sowie Probanden mit einer depressiven Störung ohne komorbide GAS wandten wir ein zweistufiges Vorgehen an: in einem ersten Schritt sollten Patienten mit einer Störung von den gesunden Kontrollprobanden unterschieden werden, im zweiten Schritt dann diejenigen Patienten mit einer GAS (mit und ohne komorbide Depression) von denjenigen ohne GAS [Hilbert et al., 2017]. In einem multimodalen Ansatz wurden Symptom-schwerefragebögen, anatomische MRT-Scans sowie phasische Kortisolausschüttung über den Verlauf eines Stressors (MRT-Untersuchung) als Prädiktoren berücksichtigt. Wie erwartet zeigte sich, dass Probanden mit einer Störung per Symptomfragebogen leicht von gesunden Kontrollen unterschieden werden konnten (Abb. 2). Demgegenüber eignete sich Symptomschwere aber nicht, um die Probandengruppen mit Störungen voneinander zu trennen. Interessanterweise gelang dies jedoch signifikant mit den biologischen Maßen (Neuroanatomie und Kortisolausschüttung). Mit einer Kombination aller zur Verfügung stehenden Daten gelang für beide Klassifikationen eine überzufällig genaue Zuordnung [Hilbert et al., 2017]. Die Pilotstudie demonstriert gut den potentiellen

inkrementellen Nutzen von zusätzlichen Datenmodalitäten für die Unterstützung klinischer Entscheidungen durch prädiktive Analytik.

Die Prädiktion von Risikoverläufen ist von den hier vorgestellten Fragestellungen wohl diejenige, für die bisher die wenigsten Studien vorliegen. Klinisch interessant wäre hier beispielsweise die Frage, welche Mitglieder einer Risikopopulation in einem vorgegebenen Intervall zu einer Volldiagnose übergehen, um mit dieser Information Prävention und Monitoring genauer steuern zu können. Erste vielversprechende Publikationen gibt es hier für den Bereich der PTBS. So konnten Gong et al. [2014b] Überlebende eines Erdbebens, die in der Folge eine PTBS entwickelten, von Überlebenden ohne PTBS mit einer Genauigkeit von 67% anhand von anatomischen MRT-Daten unterscheiden. Mittels Resting-State MRT-Daten konnte zudem die Schwere der PTBS-Symptome in Überlebenden mit und ohne PTBS signifikant vorhergesagt werden [Korrelation von $r = 0,32$ zwischen vorhergesagten und tatsächlichen Symptomwerten; Gong et al., 2014a]. Beide Untersuchungen sind allerdings insofern noch limitiert, als die Probanden erst untersucht wurden, als bereits eine PTBS Diagnose bestand oder nicht bestand. Es handelt sich also noch nicht um eine a priori getroffene Vorhersage über die Zeit hinweg. Dies leistet jedoch die Studie von Galatzer-Levy et al. [2014]. Mit Informationen, die in den ersten Tagen nach einem Trauma erhoben wurden, wurde das Vorliegen einer nicht-remittierten PTBS 15 Monate später vorhergesagt. Mittels der Merkmale des traumatischen Ereignisses, der darauffolgenden Versorgung durch den Notdienst und erster Symptome in den Tagen nach dem Ereignis in nahezu

1'000 Probanden wird eine area under the curve (AUC) von 0,77 für die Prädiktion erreicht [Galatzer-Levy et al., 2014]. In einer weiteren Studie mit über 20'000 Probanden in Chile konnten Interviewdaten, die etwa 3 Monate vor einem schweren Erdbeben erhoben wurden, zur Vorhersage von Symptomen auf dem Schweregrad vergleichbar einer PTBS etwa 3 Monate nach dem Erdbeben genutzt werden, wobei eine AUC von 0,79 erreicht wurde [Rosellini et al., 2018]. Einen aktuellen Überblick über Prädiktionsstudien zur PTBS bieten Schultebraucks und Galatzer-Levy [2019]. Weitere Studien zur Prädiktion des Onsets von Psychosen oder von einer Major Depression sind ebenfalls verfügbar [z.B. Koutsouleris et al., 2009; Mechelli et al., 2011; Foland-Ross et al., 2015].

Für die Vorhersage von Behandlungsverläufen und -ergebnissen liegen ebenfalls eine Reihe von Untersuchungen vor, viele davon aus dem Bereich der KVT für verschiedene Störungen wie Zwangserkrankungen [Askland et al., 2015; Reggente et al., 2018], soziale Angststörungen [Månsson et al., 2015; Whitfield-Gabrieli et al., 2016] oder Panikstörungen mit bzw. ohne Agoraphobie [Hahn et al., 2015]. Daneben gibt es ebenfalls Untersuchungen zur Vorhersage der Ergebnisse medikamentöser Behandlungen, beispielsweise für Antidepressiva [Chekroud et al., 2016; Iniesta et al., 2016], oder zum Vergleich verschiedener Behandlungsoptionen, etwa für KVT versus eye movement desensitization and reprocessing [EMDR; Deisenhofer et al., 2018] bei Patienten mit PTBS oder KVT versus psychodynamische Psychotherapie bei Patienten mit depressiven Störungen [Cohen et al., 2020]. Diese Studien nutzten teils soziodemographische und klinische [Chekroud et al., 2016; Iniesta et al., 2016; Deisenhofer et al., 2018; Cohen et al., 2020], teils (f)MRT-Daten [Hahn et al., 2015; Månsson et al., 2015; Reggente et al., 2018] als Prädiktoren in ihren Vorhersagen. In den Studien, die eine für den individuellen Patienten besser geeignete Behandlung vorhersagten, zeigten sich meist tatsächlich günstigere Ergebnisse in dimensionalen Symptomschweregraden für diejenigen Patienten, die die als optimal vorhergesagte Behandlung erhielten. Bei denjenigen Studien, die ein kategoriales Ergebnis (z.B. Remission oder klinisch-relevante Verbesserung: ja/nein) vorhersagten, zeigten sich meist signifikant überzufällig gute Vorhersagegenauigkeiten. Interessanterweise waren die Genauigkeiten in den auf (f)MRT-Daten basierenden Studien meist besser. Jedoch waren die Stichprobengrößen in diesen Fällen oftmals erheblich kleiner, was mutmaßlich mit einer Überschätzung der Vorhersagegüte einhergeht [Varoquaux, 2017]. Anders als klinische Routedaten liegen zudem neurobiologische Daten (z.B. [f]MRT-Scans) in der Regel für Patienten mit psychischen Störungen nicht vor. Daher stellt sich aus einer praxisorientierten Perspektive die Frage, welche Vorhersagegüte mit Daten, die routinemäßig im Rahmen des diagnos-

tischen Prozesses erhoben werden, für die prädiktive Analytik erzielt werden können. In einer Untersuchung an einem naturalistischen Datensatz von über 2'000 Patienten des Zentrums für Psychotherapie der Humboldt-Universität zu Berlin (ZPHU) sagten wir eine Remission oder Non-Remission der Primärdiagnose für den individuellen Patienten vorher [Hilbert et al., 2020]. Hierbei berücksichtigten wir alle am ZPHU behandelten Diagnosen, mithin also die gesamte Breite des Indikationsspektrums für KVT, wobei der Schwerpunkt im Bereich der internalisierenden Störungen lag. Zur Vorhersage wurden nur routinemäßig erhobene soziodemographische und klinische Basisdaten aus Diagnostik und Probatorik verwendet, wie beispielsweise Primär- und komorbide Diagnosen, funktionelle Beeinträchtigungen, Symptomschweregrade oder Alter, Geschlecht und Bildungsgrad. Vergleichbar zu anderen Prädiktionsstudien mit ähnlichen Ausgangsvariablen und kleineren Stichprobengrößen fanden wir auch hier eine signifikante, überzufällig gute Vorhersagegüte, die aber insgesamt nur von moderater Größe war. Die Beschränkung auf homogenere Diagnosegruppen innerhalb des Datensatzes führte dabei interessanterweise nicht zu einer Verbesserung der Vorhersagegüte. Aufgrund ihrer hohen externalen Validität liefert die Untersuchung eine präzise Schätzung der Möglichkeiten zur Prädiktion von KVT-Behandlungserfolgen unter naturalistischen Bedingungen zur aktuellen Zeit, basierend auf routinemäßig vorliegenden Informationen [Hilbert et al., 2020]. Es wäre zu überlegen, unter welchen Bedingungen bzw. mit welchen zusätzlichen Daten eine klinisch bedeutsame Prädiktionsgenauigkeit erzielt werden könnte.

Limitationen und Ausblick

Die hier exemplarisch vorgestellten Studien demonstrieren, dass auch klinische Psychologen und Psychotherapeuten und deren Patienten von mittels prädiktiver Analytik bereitgestellten Markern profitieren könnten. Diagnosen könnten potentiell zuverlässiger, Betroffene mit einem hohen Risiko einer Exazerbation eher identifiziert und Indikationsentscheidungen stärker evidenzbasiert getroffen werden. Ähnlich wie bei somatischen Erkrankungen und Verläufen dienten derartige Marker als neue Informationsquelle zur Unterstützung klinisch relevanter Entscheidungen. Der Behandler selbst bleibt der Entscheidungsträger, kann aber zusätzliche Informationen zu Rate ziehen, die seine Entscheidung unterstützen. Gleichzeitig wird deutlich, dass bis zur tatsächlichen Einführung solcher Marker in unseren klinischen Alltag noch einige Hürden zu nehmen sind, von denen nicht alle im engeren Sinne wissenschaftlich sind.

Im Vordergrund steht erst einmal die weitere Modellentwicklung und Prüfung. Bisherige Studien sind vielversprechend, die berichteten Modelle sind im überwiegenden Fall jedoch noch nicht leistungsfähig genug für die klinische Praxis. Mitunter sind die untersuchten Fragestellungen zu Demonstrationszwecken noch trivial. Um tatsächlich den Sprung in die klinische Anwendung zu schaffen, müsste ein potentieller Marker für eine klinisch hoch relevante Fragestellung vorliegen und dabei einen bedeutsamen Mehrwert zur bisherigen Praxis liefern. Zudem müsste dieser Mehrwert den zusätzlichen Aufwand übertreffen. Momentan wird die Frage der Leistungsfähigkeit solcher Marker vor allem vor dem Hintergrund der genutzten Datenmodalitäten diskutiert. Einerseits wird davon ausgegangen, dass die Integration verschiedener Modalitäten, wie beispielsweise klinische Informationen, hirnelektrische sowie genetische Daten, vermutlich zu einer verbesserten Vorhersagegüte führt. Diese Annahme scheint plausibel, erhöht allerdings gleichzeitig den Aufwand zur Erhebung der Marker beträchtlich. Insbesondere mit Blick auf die Effizienz scheint daher die Nutzung schon jetzt erhobener Routinedaten und per Smartphone aktiv (z.B. EMA) sowie passiv (Bewegungsmuster, Aktivität, Schlafzeiten usw.) gesammelter Daten empfehlenswert zu sein. Neben der Datenmodalität selbst sind noch weitere Aspekte für eine möglichst hohe Vorhersagegüte wesentlich. Zum einen ist der mitunter unterschätzte Aspekt der Reliabilität der Prädiktoren sowie der abhängigen Variable relevant. Wie immer begrenzt die Reliabilität letztendlich die Validität. Selbst ein noch so guter Lernalgorithmus wird keine hohe Vorhersagegenauigkeit aufweisen, wenn Prädiktor und Ergebnis durch starke Messfehler verzerrt sind. Zum anderen stellt sich jenseits der Datenmodalität die Frage, welche Variablen inhaltlich für Modelle der prädiktiven Analytik genutzt werden sollen. Mutmaßlich sind Variablen umso besser zur Prädiktion geeignet, je genauer sie die für die jeweilige Fragestellung relevanten Mechanismen abbilden. Zur Vorhersage des Therapieergebnisses besitzen also mutmaßlich diejenigen Variablen einen hohen Informationsgehalt, die Mechanismen der Aufrechterhaltung von Psychopathologie einerseits und Veränderungsmechanismen andererseits gut abbilden. Hier könnten also beispielsweise repetitives negatives Denken [Wahl et al., 2019] oder Emotionsregulationskompetenzen [Fernandez et al., 2016] relevanter sein als soziodemographische Informationen oder Symptomschwere, wie sie in vielen Studien untersucht werden. Auch basale Funktionen menschlichen Erlebens und Verhaltens, wie sie beispielsweise in den Research Domain Criteria (RDoC) [Insel et al., 2010; Cuthbert und Insel, 2013; Walter, 2017] beschrieben werden, könnten hier einen Mehrwert bieten. Nicht zuletzt stellt sich zum gegenwärtigen Zeitpunkt die Frage nach der externalen Validität der un-

tersuchten Stichproben und Algorithmen. Die meisten Studien zur prädiktiven Analytik basieren auf homogenen Stichproben, die im Rahmen klinischer Studien gewonnen wurden. Diese Patienten repräsentieren aber nicht ausreichend die heterogenen Patientengruppen der klinischen Praxis. Zukünftige Forschungsansätze, die die Implementation prädiktiver Analytik für die klinische Praxis vorantreiben wollen, sollten also entsprechend heterogene Stichproben rekrutieren, die der Versorgungswirklichkeit entsprechen. Multizentrische Forschungsprojekte wie das KODAP-Projekt [Koordination der Datenerhebung und -auswertung an Hochschul- und Ausbildungsambulanzen für Psychotherapie; Velten et al., 2017] sind hilfreiche Initiativen zum Aufbau von Infrastrukturen, die die Generierung umfangreicher, ökologisch valider Datensätze aus der klinischen Versorgung als Grundlage für Ansätze der prädiktiven Analytik ermöglichen.

Zur Prüfung potentieller Marker liegt eine exzellente Übersicht von Woo et al. [2017] vor. Nach der Entwicklung eines Markers und der initialen Demonstration einer guten Vorhersage für eine relevante Fragestellung sollten mehrere stetig anspruchsvollere Prüfungen erfolgen. Zuerst erfolgt eine direkte Replikation an einem neuen, aber hinsichtlich des Settings möglichst vergleichbaren Datensatz. Ist diese Hürde genommen, schließen sich weitere Überprüfungen an, zuerst in anderen Einrichtungen, die sich nun auch hinsichtlich des Settings und damit zusammenhängender Rahmenvariablen wie Alter oder Diagnosespektrum der Patienten unterscheiden, zuletzt dann auf Bevölkerungsebene [Woo et al., 2017]. Dabei wird nicht nur geprüft, ob der Marker unter immer naturalistischeren Bedingungen weiterhin zufriedenstellende Güte aufweist, vielmehr können durch die Veränderung der Rahmenvariablen nun auch die Robustheit und Grenzen des Modells abgeschätzt werden [Woo et al., 2017].

Während im Feld aktuell neue Modelle exploriert und bereits bestehende, vielversprechende Modelle optimiert und weiter getestet werden, stellt sich zudem die Frage, welche Vorhersagegüte letztlich überhaupt erreichbar ist. Die absolut überwiegende Mehrheit der bisherigen Studien zur Einzelfallprädiktion von Psychotherapieerfolgen nutzt beispielsweise ausschließlich Patientenvariablen zur Prädiktion. Neben Patientenvariablen spielen allerdings auch Therapeutenvariablen sowie Therapie(Interventions-)variablen eine Rolle für den Gesamterfolg einer Therapie [Heinonen et al., 2012; Wampold und Imel, 2015]. Ein Modell, das ausschließlich den Patienten in den Blick nimmt, kann daher womöglich nicht die gesamte Varianz möglicher Therapieausgänge erklären. Mit dieser Einschränkung kann allerdings auf mindestens zwei Arten umgegangen werden. Zum einen können Einflussfaktoren außerhalb des Patienten als Settingvariablen betrachtet werden, unter denen das Modell gilt. Ein Modell könnte

also beispielsweise explizit nur für depressive Patienten, die in einem universitären Ausbildungssetting behandelt werden, vorhersagen, ob diese von Verhaltensaktivierung profitieren werden. Hier wird also nach Patienten gesucht, die möglichst gut zu einer vorgegebenen Konfiguration von Therapeuten und Interventionsvariablen passen. Andererseits könnten Therapeuten und Interventionsvariablen explizit mit einbezogen werden. Für Patienten, die im britischen Improving Access to Psychological Therapies-Programm [IAPT; Clark et al., 2009] neu aufgenommen werden, sollen ab nächstem Jahr umfangreiche Erfolgsvorhersagen zur Verfügung stehen, die genau dies tun [May, 2019]. Hier werden einerseits wiederum Patientenvariablen zu einer individuellen Erfolgsprädiktion pro Patient verrechnet, andererseits basierend auf früheren Behandlungen aber auch Erfolgsprofile pro individuellem Behandler erstellt. Zudem sind individuelle Behandler bestimmten Behandlungstypen (z.B. KVT, EMDR, Familienberatung) zugeordnet. Diese Informationen werden nun zu einer individuellen Erfolgswahrscheinlichkeit pro Patient und Behandlungsalternative zusammengeführt [May, 2019]. Interessant ist dabei auch, dass der Patient neben der Erfolgswahrscheinlichkeit für die aussichtsreichsten Alternativen gleichzeitig Informationen über deren Kosten und Verfügbarkeit bekommen soll.

Einen ersten kurzen Überblick über weitere Hürden über die reine wissenschaftliche Fundierung hinaus bieten Singh und Rose [2009]. Darunter fallen beispielsweise Fragen zum Datenschutz oder zur Legitimation der Verwendung von Patientendaten zur weiteren Verbesserung der Vorhersagealgorithmen, aber auch zur Transparenz und Kommunikation von Modellentwicklungen und -grenzen oder zur Diskriminierungsfreiheit. So wird die Möglichkeit, Routinedaten der Krankenkassen auf Antrag für Forschungszwecke zur Verfügung zu stellen, aktuell im Rahmen des Digitalen-Versorgung-Gesetzes [Gesetzentwurf der Bundesregierung, 2019] kontrovers diskutiert. Neben dem reinen Nutzen derartiger Prädiktionsmodelle im Sinne einer hohen Vorhersageleistung wird die Klärung dieser Fragen zentral sein, um sowohl Patienten als auch Kliniker zu überzeugen und Akzeptanz für Vorhersagealgorithmen zur Unterstützung klinischer Entscheidungen herzustellen, selbst wenn die zu-

grundlegenden Modelle mutmaßlich so komplex sind, dass sie von Betroffenen und Anwendern im Einzelnen nicht mehr vollumfänglich verstanden werden können. Zuletzt müsste die Versorgungslandschaft einen solchen Ansatz tatsächlich auch zulassen. Die besten Prädiktoren nützten wenig, wenn beispielsweise aufgrund von begrenzten Behandlungskapazitäten gar keine Freiheitsgrade für eine differentielle Indikation (z.B. Pharmako- oder psychotherapeutische Behandlung) bestehen.

Insgesamt lässt sich also feststellen, dass psychische Störungen komplex und multifaktoriell determiniert sind, was die Entwicklung von Markern zur Unterstützung klinischer Entscheidungen im Vergleich zur somatischen Medizin deutlich erschwert und verzögert hat. Die prädiktive Analytik und die Methoden des maschinellen Lernens bieten eine Möglichkeit, aus einer Vielzahl miteinander interagierender Variablen eine Vorhersage für den individuellen Patienten abzuleiten. Erste Studien haben diese Methode für diagnostische oder differentialdiagnostische Zwecke, zur Vorhersage von Risikoverläufen sowie zur Vorhersage von Behandlungsergebnissen genutzt und bereits vielversprechende Ergebnisse erzielt. Bis zum Einsatz in der klinischen Praxis wird das Gros der wissenschaftlichen Arbeit auf der Optimierung und Prüfung solcher Marker für heterogene Patientengruppen der klinischen Versorgungsrealität liegen. Zudem besteht weiterer ethischer und gesellschaftlicher Diskussionsbedarf, wobei eine enge Zusammenarbeit mit Vertretern von Praktikern und Betroffenen wünschenswert ist. Gelingt dies, könnte die prädiktive Analytik einen substantiellen Beitrag zu Diagnostik und Therapieplanung leisten und darüber hinaus die Behandlungsergebnisse und deren Kosteneffizienz weiter verbessern.

Statement of Ethics

Für diesen Überblicksartikel war keine Begutachtung durch eine Ethikkommission notwendig.

Disclosure Statement

Es bestehen keine Interessenkonflikte.

Literatur

- APA. Diagnostic and statistical manual of mental disorders. 4th ed., text rev. Washington (DC): American Psychiatric Association; 2000.
- APA. Diagnostic and statistical manual of mental disorders. 5th ed. Washington (DC): American Psychiatric Association; 2013.
- Askland KD, Garnaat S, Sibrava NJ, Boisseau CL, Strong D, Mancebo M, et al. Prediction of remission in obsessive compulsive disorder using a novel machine learning strategy. *Int J Methods Psychiatr Res.* 2015 Jun;24(2):156–69.
- Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 2001 Mar;69(3):89–95.
- Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
- Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods.* 2018 Apr;15(4):233–4.

- Bzdok D, Krzywinski M, Altman N. Points of Significance: Machine learning: a primer. *Nat Methods*. 2017 Nov;14(12):1119–20.
- Callego J, Stanley MA, Greisinger A, Wehmanen O, Johnson M, Novy D, et al. Generalized anxiety disorder in older medical patients: diagnostic recognition, mental health management and service utilization. *J Clin Psychol Med Settings*. 2009 Jun;16(2):178–85.
- Carter RM, Wittchen HU, Pfister H, Kessler RC. One-year prevalence of subthreshold and threshold DSM-IV generalized anxiety disorder in a nationally representative sample. *Depress Anxiety*. 2001;13(2):78–88.
- Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry*. 2016 Mar;3(3):243–50.
- Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B. Improving access to psychological therapy: initial evaluation of two UK demonstration sites. *Behav Res Ther*. 2009 Nov;47(11):910–20.
- Cohen ZD, Kim TT, Van HL, Dekker JJ, Driessen E. A demonstration of a multi-method variable selection approach for treatment selection: recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychother Res*. 2020 Feb;30(2):137–150.
- Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med*. 2013 May;11(1):126.
- Deisenhofer AK, Delgadillo J, Rubel JA, Böhnke JR, Zimmermann D, Schwartz B, et al. Individual treatment selection for patients with posttraumatic stress disorder. *Depress Anxiety*. 2018 Jun;35(6):541–50.
- Engel GL. The need for a new medical model: a challenge for biomedicine. *Science*. 1977 Apr;196(4286):129–36.
- Gesetzentwurf der Bundesregierung. Entwurf eines Gesetzes für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz – DVG). 2019. Available from: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/D/Digitale-Versorgung-Gesetz_DVG_Kabinett.pdf
- Fernandez KC, Jazaieri H, Gross JJ. Emotion Regulation: A Transdiagnostic Perspective on a New RDoC Domain. *Cognit Ther Res*. 2016 Jun;40(3):426–40.
- Foland-Ross LC, Sacchet MD, Prasad G, Gilbert B, Thompson PM, Gotlib IH. Cortical thickness predicts the first onset of major depression in adolescence. *Int J Dev Neurosci*. 2015 Nov;46:125–31.
- Galatzer-Levy IR, Karstoft KI, Statnikov A, Shalev AY. Quantitative forecasting of PTSD from early trauma responses: a Machine Learning application. *J Psychiatr Res*. 2014 Dec;59:68–76.
- Gong Q, Li L, Du M, Pettersson-Yeo W, Crossley N, Yang X, et al. Quantitative prediction of individual psychopathology in trauma survivors using resting-state fMRI. *Neuropsychopharmacology*. 2014a Feb;39(3):681–7.
- Gong Q, Li L, Tognin S, Wu Q, Pettersson-Yeo W, Lui S, et al. Using structural neuroanatomy to identify trauma survivors with and without post-traumatic stress disorder at the individual level. *Psychol Med*. 2014b Jan;44(1):195–203.
- Green CA, Perrin NA, Leo MC, Janoff SL, Yarbrough BJ, Paulson RI. Recovery from serious mental illness: trajectories, characteristics, and the role of mental health care. *Psychiatr Serv*. 2013 Dec;64(12):1203–10.
- Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med*. 2019 Jan;25(1):60–4.
- Hahn T, Kircher T, Straube B, Wittchen HU, Konrad C, Ströhle A, et al. Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiatry*. 2015 Jan;72(1):68–74.
- Hahn T, Marquand AF, Ehls AC, Dresler T, Kittel-Schneider S, Jarczok TA, et al. Integrating neurobiological markers of depression. *Arch Gen Psychiatry*. 2011 Apr;68(4):361–8.
- Heinonen E, Lindfors O, Laaksonen MA, Knekt P. Therapists' professional and personal characteristics as predictors of outcome in short- and long-term psychotherapy. *J Affect Disord*. 2012 May;138(3):301–12.
- Hilbert K, Lueken U, Muehlhan M, Beesdo-Baum K. Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study. *Brain Behav*. 2017 Feb;7(3):e00633.
- Hilbert K, Kunas SL, Lueken U, Kathmann N, Fydrich T, Fehm L. Predicting cognitive behavioral therapy outcome in the outpatient sector based on clinical routine data: a machine learning approach. *Behav Res Ther*. 2020 Jan;124:103530.
- Iniesta R, Malki K, Maier W, Rietschel M, Mors O, Hauser J, et al. Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *J Psychiatr Res*. 2016 Jul;78:94–102.
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010 Jul;167(7):748–51.
- Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005 Jun;62(6):617–27.
- Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med*. 2018 Nov;24(11):1716–20.
- Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: A review of classification and combining techniques. *Artif Intell Rev*. 2006;26(3):159–90.
- Kotsiantis SB. Supervised machine learning: A review of classification techniques. *Informatica*. 2007;31(3):249–68.
- Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry*. 2009 Jul;66(7):700–12.
- Lueken U, Hahn T. Entwicklung von Vorhersagemodellen am Beispiel der Angststörungen. *Neurotransmitter (Houst)*. 2017;28(9):27–33.
- Lueken U, Hilbert K, Stolyar V, Maslowski NI, Beesdo-Baum K, Wittchen HU. Neural substrates of defensive reactivity in two subtypes of specific phobia. *Soc Cogn Affect Neurosci*. 2014 Nov;9(11):1668–75.
- Månsson KN, Frick A, Boraxbekk CJ, Marquand AF, Williams SC, Carlbring P, et al. Predicting long-term outcome of Internet-delivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning. *Transl Psychiatry*. 2015 Mar;5(3):e530.
- May C. The Application of Standardised Digital Assessments and Machine Learning for the Recommendation of Optimal Treatment Plans. Technical demonstration presented at the 9th World Congress of Behavioural and Cognitive Therapies, Berlin, Germany. 2019 July 18.
- Mechelli A, Riecher-Rössler A, Meisenzahl EM, Tognin S, Wood SJ, Borgwardt SJ, et al. Neuroanatomical abnormalities that predate the onset of psychosis: a multicenter study. *Arch Gen Psychiatry*. 2011 May;68(5):489–95.
- Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning. 2nd ed. Cambridge (MA): The MIT Press; 2018.
- Munk-Jørgensen P, Allgulander C, Dahl AA, Foldager L, Holm M, Rasmussen I, et al. Prevalence of generalized anxiety disorder in general practice in Denmark, Finland, Norway, and Sweden. *Psychiatr Serv*. 2006 Dec;57(12):1738–44.
- Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*. 2012 Apr;36(4):1140–52.
- Ozomaro U, Wahlestedt C, Nemeroff CB. Personalized medicine in psychiatry: problems and promises. *BMC Med*. 2013 May;11(1):132.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *JMLR*. 2011;12:2825–30.
- Reggente N, Moody TD, Morfini F, Sheen C, Rissman J, O'Neill J, et al. Multivariate resting-state functional connectivity predicts response to cognitive behavioral therapy in obsessive-compulsive disorder. *Proc Natl Acad Sci USA*. 2018 Feb;115(9):2222–7.
- Rosellini AJ, Dussaillant F, Zubizarreta JR, Kessler RC, Rose S. Predicting posttraumatic stress disorder following a natural disaster. *J Psychiatr Res*. 2018 Jan;96:15–22.
- Schultebras K, Galatzer-Levy IR. Machine Learning for Prediction of Posttraumatic Stress and Resilience Following Trauma: An Overview of Basic Concepts and Recent Advances. *J Trauma Stress*. 2019 Apr;32(2):215–25.

- Shim M, Hwang HJ, Kim DW, Lee SH, Im CH. Machine-learning-based diagnosis of schizophrenia using combined sensor-level and source-level EEG features. *Schizophr Res*. 2016 Oct;176(2-3):314–9.
- Singh I, Rose N. Biomarkers in psychiatry. *Nature*. 2009 Jul;460(7252):202–7.
- Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manage*. 2009;45(4):427–37.
- Strimbu K, Tavel JA. What are biomarkers? *Curr Opin HIV AIDS*. 2010 Nov;5(6):463–6.
- van Breda W, Bremer V, Becker D, Hoogendoorn M, Funk B, Ruwaard J, et al. Predicting therapy success for treatment as usual and blended treatment in the domain of depression. *Internet Interv*. 2018 Jun;12:100–4.
- Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. *Neuroimage*. 2017 June;180(Pt A):68–77.
- Velten J, Margraf J, Benecke C, Berking M, In-Albon T, Lincoln T, et al. Methodenpapier zur Koordination der Datenerhebung und -auswertung an Hochschul- und Ausbildungsambulanzen für Psychotherapie (KODAP). *Z Klin Psychol Psychother*. 2017;46(3):169–75.
- Venkatasubramanian G, Keshavan MS. Biomarkers in Psychiatry - A Critique. *Ann Neurosci*. 2016 Mar;23(1):3–5.
- Vermani M, Marcus M, Katzman MA. Rates of detection of mood and anxiety disorders in primary care: a descriptive, cross-sectional study. *Prim Care Companion CNS Disord*. 2011;13(2):e1–10.
- Wahl K, Ehrling T, Kley H, Lieb R, Meyer A, Kordon A, et al. Is repetitive negative thinking a transdiagnostic process? A comparison of key processes of RNT in depression, generalized anxiety disorder, obsessive-compulsive disorder, and community controls. *J Behav Ther Exp Psychiatry*. 2019 Sep;64:45–53.
- Walter H. Research Domain Criteria (RDoC) : psychiatrische Forschung als angewandte kognitive Neurowissenschaft. *Nervenarzt*. 2017 May;88(5):538–48.
- Wampold BE, Imel ZE. The great psychotherapy debate: the research evidence for what works in psychotherapy. 2nd ed. New York: Routledge; 2015. <https://doi.org/10.4324/9780203582015>.
- Whitfield-Gabrieli S, Ghosh SS, Nieto-Castanon A, Saygin Z, Doehrmann O, Chai XJ, et al. Brain connectomics predict response to treatment in social anxiety disorder. *Mol Psychiatry*. 2016 May;21(5):680–5.
- Wittchen HU, Kessler RC, Beesdo K, Krause P, Höfler M, Hoyer J. Generalized anxiety and depression in primary care: prevalence, recognition, and management. *J Clin Psychiatry*. 2002;63 Suppl 8:24–34.
- World Health Organization and International Programme on Chemical Safety. Biomarkers in risk assessment: validity and validation. World Health Organization; 2001. Available from: <https://apps.who.int/iris/handle/10665/42363>
- Woo CW, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*. 2017 Feb;20(3):365–77.
- Wu J, Liu X, Zhang X, He Z, Lv P. Master clinical medical knowledge at certificated-doctor-level with deep learning model. *Nat Commun*. 2018 Oct;9(1):4352.
- Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. *Knowl Inf Syst*. 2008;14(1):1–37.
- Yarkoni T, Westfall J. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect Psychol Sci*. 2017 Nov;12(6):1100–22.
- Zbozinek TD, Rose RD, Wolitzky-Taylor KB, Sherbourne C, Sullivan G, Stein MB, et al. Diagnostic overlap of generalized anxiety disorder and major depressive disorder in a primary care sample. *Depress Anxiety*. 2012 Dec;29(12):1065–71.
- Zuckerman M. *Vulnerability to psychopathology: A biosocial model*. Washington (DC): American Psychological Association; 1999.